Motivation

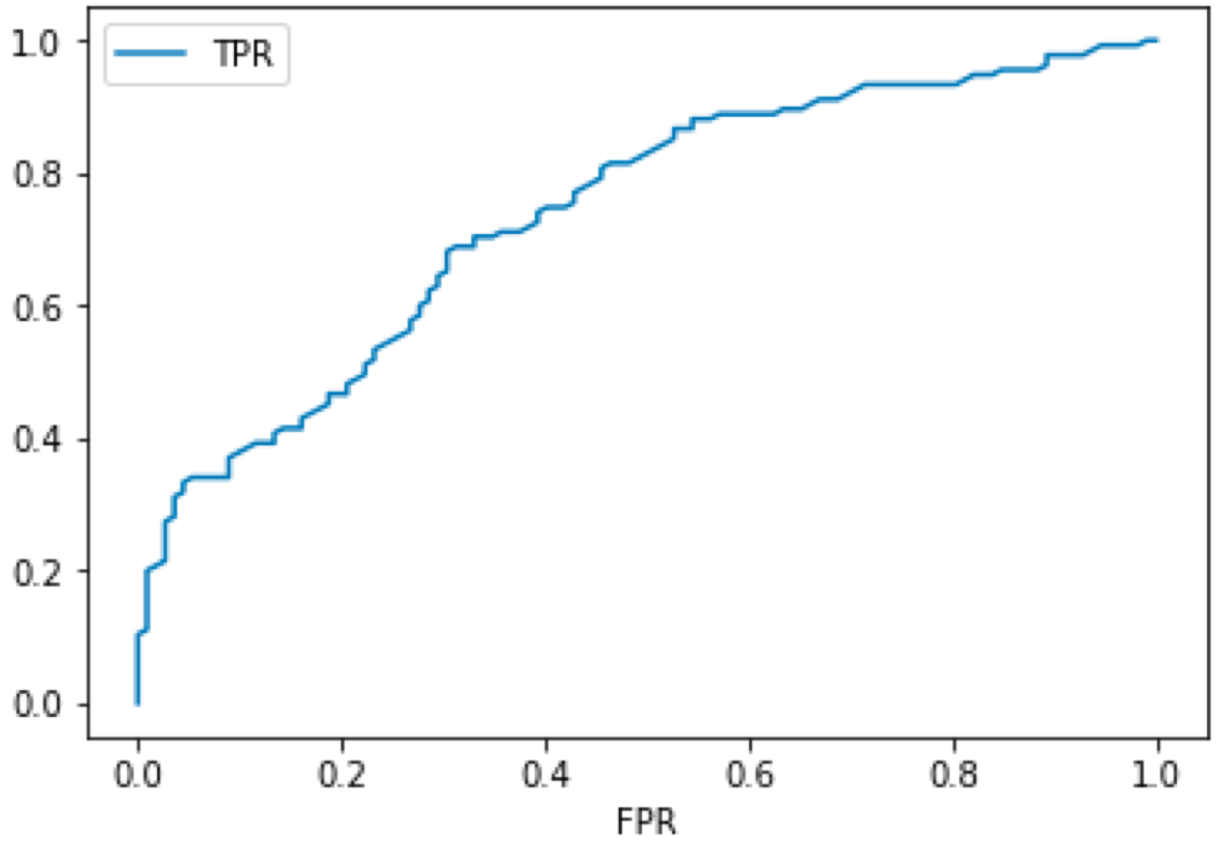
1. Heart Attack vulnerability prediction helps individuals make significant changes to their lifestyles to prevent heart attack
2. Identifying features which have impact on the heart attack vulnerability helps in designing focused plan to improve lifestyle
3. Learn Apache Spark for ML model training

Code Snippet

```
# Using VectorAssembler to vectorize multiple features

from pyspark.ml.feature import VectorAssembler

# defining custom function to convert input into vector
sqlc.registerFunction("oneElementVec", lambda d: Vectors.dense([d]), returnType=VectorUDT())

featureCols = ['age','sex','trestbps','chol']



assembler = VectorAssembler(inputCols=featureCols,outputCol="features")
trainingDF = assembler.transform(trainingDF)

#### Ref - https://stackoverflow.com/questions/32556178/create-labeledpoints-from-spark-dataframe-in-python
```

This code vectorizes multiple features using VectorAssembler and transforms dataframe

Visualisations



ROC curve with 74% AUC on test data
Heart attack classification